



# Info@starklawlibrary.org

Stark County Law Library Association

June 2002

## BEGINNERS

## What's a Bot?

The email  
address,  
**info@starklaw  
library.org**  
really works!  
**TRY IT!!**



Bots go by a  
lot of other  
names but  
basically they  
perform  
similar tasks  
in a slightly  
different  
manner.

### What is a Bot?

A **bot** is a software tool for digging through data. You give a bot directions and it bring(s) back answers. The word is short for **robot** of course, which is derived from the Czech word *robota* meaning work.<sup>1</sup>

Bots go by a lot of other names such as agents, crawlers, spiders or ants. Strictly speaking, an agent is a bot (usually utilizing artificial intelligence) that goes out on a mission usually to find information and report back while other bots operate in one place. Spiders and ants perform similar tasks in slightly different manners.



For more in-depth information on bots, what they are and how they work you can check out the following list of articles and websites.

**“Crawlers, Spiders, Robots”** at: <http://www.geocities.com/toe6000/search1.html> explains how sites are scanned for links to other sites, then the technology moves on to these other sites. More than a million sites a day can be visited with bots, building databases by indexing the contents of Web sites.



**“Robots on the Internet”** at: <http://www.wsu.edu/~gened300/internet/robot.htm> offers the following tips on maximizing the efficiency of searching these Robots:

1. Use quotation marks around a group of words
2. Use a plus (+) sign in front of a search term ("juvenile delinquency" + "neighborhood watch") will usually result in more accurate results than ("juvenile delinquency" and "neighborhood watch.")

**“BotSpot's Sweet Spot”** at: <http://www.abanet.org/tech/ltrc/site-tation/october2001.html> is the home of "... programs that are free, whimsical and technically primitive to those that are expensive, serious and technically sophisticated."<sup>2</sup> At this site you will find:

**LawCrawler** at: <http://lawcrawler.findlaw.com/> is a comprehensive SearchBot for legal information. It searches United States and individual country domains. It utilizes Google's database.

**Law Engine** at: <http://www.thelawengine.com/> is an online law resource available in a fast, well organized, easy-to-use, single page format. It uses Fast (alltheweb) for its searching.

Continued on page 2

## BEGINNERS

## What's a Bot? (Continued)

Continued from page 1

**LawGuru** offers two search components. Legal Research Page at <http://www.lawguru.com/search/lawsearch> from which you can search more than 535 legal search engines' tools using www.wais 2.5 and their Multi Resource Legal Research page at <http://www.lawguru.com/multisearch/multimenu.html>.

**Law Runner** at: <http://www.lawrunner.com> DOES use predefined intelligent agents with AltaVista™ and select legal databases in its searches.

**P.S.** They also list LawTrack that went out of business early last year!

If bots and search engines sound the same to you, it is because bots are used to create the databases from which search engines get their results. The difference is mainly one of timing. Search engines update some of their

sources daily, but others are updated monthly! Whereas you can send a bot out at any time to gather specific information of your choosing.

A few examples of other bots that could be useful are: **Webspector** at: <http://www.illumix.com/webspector.htm> <sup>3</sup> **C4U** <sup>4</sup> at: <http://www.c4u.com/>. Both track specific websites for changes. **eWatch** at: <http://www.ewatch.com/> monitors Internet publications, discussion forums, bulletin boards and electronic mailing lists for references to the company or individual you specify, <sup>5</sup> **Speechbot** at: <http://speechbot.research.compaq.com/>, sponsored by Hewlett-Packard, is the first search engine for audio visual material. <sup>6</sup> **TVEyes** at: <http://www.tveyes.com/> specializes in real-time or near real-time monitoring of people or companies from broadcast material. <sup>7</sup>

## INTERMEDIATE

## The Wayback Machine

**The Internet has become the repository for a vast amount of information but it has suffered from a lack of permanence. The Wayback Machine can help you find old information.**

The Internet has become the repository for a vast amount of information but it has suffered from a lack of permanence, in part due to the numerous failures of "dot com" businesses.

Enter The Wayback Machine at: <http://www.archive.org>.

Starting in 1996, the Internet Archive has been storing Web pages that Brewster Kahle's Alexa Internet has crawled. The Wayback Machine, launched in October of 2001 makes available to the public more



than 100 terabytes of data and is growing at a rate of 12 terabytes per month making it the largest known database in the world. <sup>11</sup>

The GOOD.

- Although basic access to the archive is by URL, it does have an advanced search form offering many limiting features. It is available at: <http://web.archive.org/collections/web/advanced.html>, or by reading the fine print at the top right side of the search form.

Continued on page 3

## INTERMEDIATE Wayback Machine (Continued)

Continued from page 2

- Although it is not text searchable, it can be browsed (expect some dead links).
- It is wonderful
  - a) when you come across the inevitable “404 not found” or similar messages,
  - b) for conducting historical research, or
  - c) tracking down sources lost due to complex URL shifting

The BAD.

- Anyone can have their site removed from the Archive by using a robots.txt file on their server.<sup>12</sup>
- The Archive will not always be available. Expect “higher than expected number of requests” or “down for maintenance” messages.

The UGLY.

- Various Internet filter programs block access to the archive because it contains the archives of “sex sites and other material that would be blocked by filters.”<sup>13</sup>

## ADVANCED Alexa Toolbar

**Alexa collects personally identifiable information and so does Amazon.com (which owns Alexa)!**

Do I Really need TWO extra toolbars on my browser? YES!

The Alexa Toolbar will provide you with:

- *Site Information* for the site you are visiting including: Contact Information, Reviews and Site Stats (if available)
- *Ranking of the Site* (which is questionable, read about it at: [http://pages.alexa.com/prod\\_serv/new.html?p=Det\\_W\\_t\\_40\\_M1](http://pages.alexa.com/prod_serv/new.html?p=Det_W_t_40_M1)) and note that:

“It doesn't say how frequently the data on site ranking is updated. It doesn't make clear if it gives a default ranking to sites not registered. Some sites appear to have a ranking of +14,000,000, I'm not sure if this means they haven't registered, or just that no-one has visited the site!”<sup>14</sup>

- Meta Search Engine
- *Wayback Machine* allowing you easy access without having to use “Favorites” and typing a long URL!

To download the Alexa Toolbar go to: [http://download.alexa.com/alexa65/startpage.html?p=dest\\_W\\_t\\_40\\_B2](http://download.alexa.com/alexa65/startpage.html?p=dest_W_t_40_B2) and click on “Install Alexa Toolbar.” You may also want to take their tour that is available at the bottom of the page

Alexa Toolbar users are advised to READ the “Privacy Policy” located at: <http://www.alexa.com/help/privacy.html>. Alexa DOES collect personally identifiable information and so does Amazon.com (which owns Alexa)! Get the full story by going to: <http://resourceshelf.freepint.com/> and scroll down to April 30th and “Alexa Meets Google.”

## FOOTNOTES

- <sup>1</sup> “What’s a bot?” internet.com Botspot. 2002 INT Media Group. 9 May 2002. <[http://www.botspot.com/bot/what\\_is\\_a\\_bot.html](http://www.botspot.com/bot/what_is_a_bot.html)>.
- <sup>2&5</sup> Thompson, Kathryn. “BotSpot's Sweet Spot.” ABA Legal Technology Resource Center. Oct 2001. ABAnetwork. 9 May 2002. <<http://www.abanet.org/tech/ltrc/site-tation/october2001.html>>.
- <sup>3</sup> “Downloads: Webspector™.” Illumix Software. 1999 Illumix Software. 9 May 2002. <<http://www.illumix.com/downloads.htm>>.
- <sup>4</sup> Calishain, Tara. “Reader Response -- Web Site Monitoring Software.” Research Buzz News: May 2-8, 2002. 9 May 2002. ResearchBuzz.com. 14 May 2002. <<http://www.researchbuzz.com/news/2002/may02may08.shtml>>.
- <sup>6</sup> “Audio Search Using Speech Recognition.” SpeechBot™. 2002 Hewlett-Packard Company. 9 May 2002. <<http://speechbot.research.compaq.com/>> cited in Virtual Acquisition Shelf and News Desk. May 9, 2002. <<http://resourceshelf.freepint.com/>>.
- <sup>7</sup> Price, Gary. “Part 2: Searchable Broadcast Transcripts--Fee Based Services Monitoring Services.” Virtual Acquisition Shelf and News Desk. 9 May 2002. Gary Price Library & Internet Research Consulting. 14 May 2002. <<http://resourceshelf.freepint.com/>>.
- <sup>8 & 9</sup> Guernsey, Lisa. “Upstart Search Engines Try to Topple Google.” Technology. 4 April 2002. The New York Times Company. 15 April 2002. <<http://college4.nytimes.com/guests/articles/2002/04/04/911647.xml>>. Cited from <<http://resourceshelf.freepint.com/>> on Thursday, 4 Apr 2002.
- <sup>10</sup> Teoma. 2002. Ask Jeeves, Inc. 16 April 2002. <<http://www.teoma.com/search.asp?t=%93blog+%2Blaw%94&qcat=1&qsrc=0&search.x=31&search.y=11>>
- <sup>11</sup> “Wayback Machine Online Press Kit: Wayback Machine Overview.” Internet Archive: Building an ‘Internet Library.’ 10 March 2001. Internet Archive. 13 May 2002. <[http://www.archive.org/wayback/press\\_kit/index.html](http://www.archive.org/wayback/press_kit/index.html)>. Cited from <<http://resourceshelf.blogspot.com/>> on 26 October 2001.
- <sup>12</sup> Notess, Greg R. “The Wayback Machine: The Web's Archive.” Online. March/April 2002. InfoToday 2002. <<http://www.infotoday.com/online/mar02/OnTheNet.htm>> cited from “TVC Alert” 5 March 2002 <<http://www.virtualchase.com/TVCAAlert/mar02/5mar02.html>>
- <sup>13</sup> “Internet Filters Block Wayback Machine.” Library Journal. 28 Mar 2002. Reed Elsevier Inc. 13 May 2002. <<http://libraryjournal.reviewsnews.com/index.asp?layout=article&articleid=CA202791&article&display=breakingNews&publication%20=libraryjournal&publication=libraryjournal>>
- <sup>14</sup> “Calculation of Alexa Rankings.” Webmaster World.com. 2 March 2002. Westhost.com. 13 May 2002. <<http://www.webmasterworld.com/forum31/300.htm>>